Self-Similarity and Points of Interest

Jasna Maver

Abstract—In this work, we present a new approach to interest point detection. Different types of features in images are detected by using a common computational concept. The proposed approach considers the total variability of local regions. The total sum of squares computed on the intensity values of a local circular region is divided into three components: between-circumferences sum of squares, between-radii sum of squares, and the remainder. These three components normalized by the total sum of squares represent three new saliency measures, namely, radial, tangential, and residual. The saliency measures are computed for regions with different radii and scale-spaces are build in this way. Local extrema in scale-space of each of the saliency measures are located. They represent features with complementary image properties: blob-like features, corner-like features, and highly textured points. Results obtained on image sets of different object classes and image sets under different types of photometric and geometric transformations show high robustness of the method to intra-class variations as well as to different photometric transformations and moderate geometric transformations and compare favourably with the results obtained by the leading interest point detectors from the literature. The proposed approach gives a rich set of highly distinctive local regions that can be used for object recognition and image matching.

Index Terms—interest point detector, self-similarity, visual attention, visual perception, linear predictors, the total sum of squares.

1 INTRODUCTION

In the past few years, many techniques have been proposed to try to locate interest points in images [12], [29], [21], [41], [35] [14], [26], [34], [36], [27], [25], [42], [43], [38], [39], [15], [10]. These locations are then used either for image matching or for extraction of meaningful entities from images that can further be applied for object or scene representation and recognition. In the papers [26] and [15] different detectors were tested on image sets under different types of photometric and geometric transformations and on image sets of different object classes. Obtained results show that there is not one detector that outperforms the others for all scene types and all types of transformations. A set of complementary detectors is needed to cope with different phenomena in images.

In this work, we present a new approach to interest point detection. Different types of features in images are detected by using a common computational concept. The proposed approach is motivated in the following way: The number of objects and scenes that an intelligent vision system needs to know and recognize is huge. Hence, every system for object recognition should pay attention to the amount of data that is needed for its representation. It makes sense then to find locations in images with local region structure that can be coded efficiently and use these locations for object representations. When a local region is self-similar, the data is redundant and we can expect that less data is needed for its representation. Locations in images with self-similar structure of a local pattern are also distinguishable from locations in their vicinity, and hence, can also be used for image matching. In this paper, we propose measuring self-similarity of local regions for interest point detection.

Symmetry is an example of self-similarity. In [1], [2], [9], symmetry of plane curves has been analyzed for describing shapes. For this purpose, different descriptive terms were introduced, as, for example: sym-ax, smoothed local symmetry, symmetry set, and midlocus. The aim was to reduce the huge amount of information carried by a shape down to a “skeleton” of crucial information, which can be more readily assimilated. Closer to our work are context-free attentional operators which tend to use radial symmetry for interest point detection. Two types of approaches can be identified; both rely on image gradient. In [31], [20], [40], symmetry maps are obtained by computing the contribution a local neighbourhood makes to the symmetry measure of a central pixel. The second approach [33], [16], [24] follows the idea of the circular Hough transform [11], [28]. Symmetry maps are obtained by computing the contribution each pixel makes to the symmetry of pixels around it. The second approach allows faster computation of symmetry maps. The best runtime, well-suited for real-time vision applications, is achieved by the fast radial symmetry transform of Loy and Zelinsky [24].

Di Gesù and Valenti [3] propose a discrete symmetry transform based on local axial moments. The method has been applied to eye detection [6], processing astronomical images [4], and as an early vision process [5]. Recursive methods [6], [30] reduce the computational load of the symmetry transform. A drawback of the applied symmetry transform is its tendency to highlight lines and regions of high texture in addition to radially symmetric features.

Kovesi [17] proposes a measure of symmetry that is based on the analysis of local frequency information;
local symmetry and anti-symmetry in images can be identified as particular arrangements of phase. Local frequency information is determined via convolution with quadrature log-Gabor filters for the full range of filter orientations and a number of scales. The proposed technique is invariant to the magnitude of the local contrast. The computational cost of the method is high.

Recent papers on symmetry are [13], [32], [7], [8], [23], [19]. In this work, the main motivation is no longer interest point detection. Instead, effort is put into developing an efficient method to detect objects with symmetry, to classify this symmetry or to extract regions with symmetry from images.

In this work, self-similarity of a local region is measured on image intensity values and is quantified by the standardized regression coefficients. Three new measures of region saliency are derived. They are invariant to the similarity group of geometric transformations and to photometric shift. Because they are based on the standardized regression coefficients, they represent precise measures and allow a scale space approach. Local extrema in scale-spaces of the proposed saliency measures represent different types of features: blob-like features, corner-like features and highly textured points.

The approach proposed in this work enhances the context-free attentional operators based on radial symmetry. Gradient-based image feature detection and extraction approaches are sensitive to variations in image illumination, blurring, and magnification. The threshold applied needs to be modified appropriately [18]. Our method is not gradient-based, quite the contrary. It introduces new entities: circumferences and radii. Values on circumferences and radii are summed together; hence, the proposed approach integrates information in images at the very first level of data processing, which makes it robust to noise. Results obtained on image sets of different object classes and image sets under different types of photometric and geometric transformations show high robustness of the method to intra-class variations as well as to different photometric transformations and moderate geometric transformations, and compare favourably with the results obtained by the leading interest point detectors from the literature. The proposed approach gives a rich set of highly distinctive local regions that can be used for object recognition and image matching.

This paper is organized as follows. Section 2 defines region self-similarity. In Section 3, we describe quantifying region self-similarity by the normalized correlation coefficient. In Section 4, we derive three measures of region saliency which allow detection of different types of features in images. Section 5 gives implementation details. In Section 6, the proposed method is tested in accordance with two protocols ([26], [15]) for evaluation of local region detectors. Section 7 provides concluding remarks and possible extensions of the proposed method.

![Fig. 1. A reflection maps the location \((r, \phi)\) to location \((r, 2\phi - \phi)\).](image)

# 2 Self-Similarity

Imagine that you have an image fragment and an image and you would like to find a local region in the image which is similar to the image fragment. One possible way of solving this problem is to move the image fragment over the image and, at each location in the image, compute the normalized correlation coefficient between the corresponding intensity values of the image fragment and the local region. By computing the normalized correlation coefficient, it is assumed that when a local region is similar to the image fragment there exists a linear relationship between the intensity values of the image fragment and the local region. When the value of the normalized correlation coefficient is close enough to one, we believe that a local region is similar to the image fragment.

A self-similar region in an image can be detected in a similar way. At each location in the image, the normalized correlation coefficient is now computed between the intensity values of a local region and the intensity values of the same geometrically transformed local region. A local region is self-similar if there exists a linear relationship:

\[
I(T(x)) = a + bI(x), \quad \forall x \in P.
\]

Here, \(P\) denotes a circular region of radius \(R\) and \(x\) is a point location in \(P\). Variable \(I(x)\) denotes the intensity value at location \(x\), and \(T\) denotes a bijective geometric transformation defined on \(P\). Here, we limit \(T\) to reflection and rotation. Both reflection and rotation, preserve distances, angles, sizes, and shapes. For the sake of simplicity, point location is represented in polar coordinates; hence, \(x = (r, \phi)\).

## 2.1 Reflection

Every reflection has a mirror line. Let the mirror line go through the centre of \(P\) and let \(\vartheta \in [0, \pi)\) denote the mirror line orientation. A reflection maps the location \((r, \phi)\) to location \((r, 2\vartheta - \phi)\) (see Fig. 1).
2.2 Rotation

Every rotation has a centre and an angle. Let the centre of the rotation be the centre of $\mathcal{P}$ and let the rotation angle $\alpha$ be one of the angles $\frac{2\pi}{n}$, where $n$ is an integer. A rotation maps the location $(r, \phi)$ to location $(r, \phi + \alpha)$.

2.3 Parameters $a$ and $b$

We have two solutions for the parameters $a$ and $b$ of Eq. 1:

$$a = 0, \quad b = 1,$$
$$a = I(x) + I(T(x)), \quad b = -1.$$

A reflection $T$ with mirror line orientation $\vartheta$ maps the location $(r, \phi)$ to $(r, 2\vartheta - \phi)$ and location $(r, 2\vartheta - \phi)$ to $(r, \phi)$. Hence,

$$I(r, 2\vartheta - \phi) = a + bI(r, \phi),$$
$$I(r, \phi) = a + bI(r, 2\vartheta - \phi).$$

Parameters (2) solve the above system. When $T$ is a rotation, the derivation of solutions for parameters $a$ and $b$ is not straightforward; therefore, we give it in the Appendix. The line with parameters $a = 0$, $b = 1$ (see Fig. 2) represents regions where $I(T(x)) = I(x)$. These regions have symmetry (see Figs. 3(a) and 3(c) for examples). The line with parameters $b = -1$ and $a = I(x) + I(T(x))$ represents regions where the sum of intensity values at corresponding locations, that is, $I(x) + I(T(x))$, is the same value for all corresponding pairs. In this case, we say that region has anti-symmetry (see Figs. 3(b) and 3(d) for examples). Different values for $a$ give different lines. The lines corresponding to symmetry and antisymmetry are orthogonal and intersect. A point that belongs to both lines represents a region of constant intensity value. In this case, only one pair of values $(I(x), I(T(x)))$ is defined. A plotting $I(T(x))$ versus $I(x)$ can define only a point, not a line. Equation (1) is underdetermined.

3 Quantifying Region Self-Similarity

On real data, (1) can hardly be fulfilled for all points of $\mathcal{P}$ (see Figs. 3(e) and 4). Nevertheless, we can measure the strength of the linear relationship (1) by the normalized correlation coefficient:

$$ncc(P, T) = \frac{\sum_i (I(x_i) - \bar{T})(I(T(x_i)) - \bar{T})}{\sqrt{\sum_i (I(x_i) - \bar{T})^2}(\sum_i (I(T(x_i)) - \bar{T})^2)} = \sum_i (I(x_i) - \bar{T})(I(T(x_i)) - \bar{T}) \sum_i (I(x_i) - \bar{T})^2. \quad (3)$$

Here, $i$ counts all points of $\mathcal{P}$ and $T$ represents the average intensity value of points of $\mathcal{P}$. The value $ncc = 1$ is obtain for symmetry, while value $ncc = -1$ for anti-symmetry. Since the sums of squares $\sum_i (I(x_i) - \bar{T})^2$ and $\sum_i (I(T(x_i)) - \bar{T})^2$ are identical, the normalized correlation coefficient computed by (3) is the standardized regression coefficient [37]. It can be used for quantitative comparison of self-similarities computed at different locations in the image as well as self-similarities computed for regions with different radii $R$, that is, for different scale parameters of a multi-scale representation.

4 Location Saliency

At a given location, the normalized correlation coefficients (3) can be computed for different mirror line orientations or different angles of rotation. All give information of region self-similarity. We propose to use the average normalized correlation coefficient computed over all orientations of the mirror line at a given location (see Fig. 5), or over all angles of rotation as a measure of location saliency.

Computation of (3) for different mirror line orientations or angles of rotation greatly simplifies by transforming local image $\mathcal{P}$ to polar coordinates. Transformation of data from a Cartesian to polar coordinate system...
introduces a new sampling of data. Let the sampling intervals for \( r \) and \( \phi \) be \( \Delta r = \frac{R}{M} \) and \( \Delta \phi = \frac{2\pi}{N} \), respectively. A circular region of radius \( R \) is transformed to a rectangle \( P(M,N) \) with \( M \times N \) values, namely, \( I(r_m, \phi_n) \), where \( r_m = m\Delta r \) for \( m = 0, \ldots, M - 1 \) and \( \phi_n = n\Delta \phi \) for \( n = 0, \ldots, N - 1 \) (see Fig. 7). The area of a region belonging to a sampled point grows with \( \Delta r \times \Delta \phi \). A transformation of \( P \) to \( P \) corresponds to weighting of points of \( P \) by a factor \( \frac{1}{2} \). This weighting acts as a low pass filter and it normalizes circumferences to the same length, and hence, makes measures computed on circumferences comparable. It turns out that the proposed saliency measures are represented by the values computed on circumferences. We prefer using terms circumferences and radii, when having in mind rows and columns of \( P \).

### 4.1 Three Saliency Measures

#### 4.1.1 Radial Saliency

First, we compute radial saliency for reflection. Once the transformation from Cartesian to polar coordinates is done, one needs to know how many mirror line orientations are needed for computing the normalized correlation coefficient. It must be computed for all possible different sets of corresponding couples when the mapping \( T \) is a reflection. Corresponding couples can only be formed among points lying on the same circumference; hence, intensity value \( I(r_m, \phi_n) \) can form corresponding couples only with values \( I(r_m, \phi_0), I(r_m, \phi_1), \ldots, I(r_m, \phi_{N-1}) \). The symmetry couples \((I(r_m, \phi_n), I(r_m, \phi_i))\) and \((I(r_m, \phi_n), I(r_m, \phi_{i+1}))\) are formed for the mirror line orientations \( \frac{\phi_n + \phi_i}{2} \) and \( \frac{\phi_n + \phi_{i+1}}{2} \), respectively. They are separated by an angle:

\[
\frac{(\phi_n + \phi_{i+1}) - (\phi_n + \phi_i)}{2} = \frac{\phi_{i+1} - \phi_i}{2} = \frac{\Delta \phi}{2}.
\]

The mirror line orientation \( \vartheta + \pi \) gives the same symmetry couples as \( \vartheta \) and therefore does not give us any new information. \( N \) samples on a circumference require \( N \) orientations of the mirror line separated by an angle \( \Delta \vartheta = \frac{2\pi}{N} \). Let \( \phi_i = i \times \Delta \vartheta \). The set of the mirror line orientations is: \( \{ \vartheta_0, \vartheta_1, \vartheta_2, \ldots, \vartheta_{N-1} \} \). The region radial saliency when \( T \) is a reflection is computed as

\[
S_{r_{\text{refl}}}(P) = \frac{1}{N} \sum_{i=0}^{N-1} ncc(P, T_{\vartheta_i})
\]

\[
= \frac{1}{N V_P} \left( \sum_{i=0}^{N-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(r_m, \phi_n)I(r_m, 2\vartheta_i - \phi_n) \right) - \frac{1}{M} \left( \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I(r_m, \phi_n) \right)^2
\]

\[
= \frac{1}{N V_P} \left( \sum_{m=0}^{M-1} \left( \sum_{n=0}^{N-1} I(r_m, \phi_n) \right)^2 \right) - \frac{1}{M} \left( \sum_{m=0}^{M-1} \left( \sum_{n=0}^{N-1} I(r_m, \phi_n) \right)^2 \right).
\]

Here, \( V_P \) denotes the total sum of squares computed for the intensity values of \( P \), that is, \( V_P = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I(r_m, \phi_n) - T)^2 \) with \( T = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(r_m, \phi_n)}{N \times M} \). Let \( C_m = \sum_{n=0}^{N-1} I(r_m, \phi_n) \) be the sum of intensity values on the circumference with radius \( r_m \) and let \( \overline{C} = \frac{1}{N \times M} \sum_{m=0}^{M-1} C_m \). Then,

\[
S_{r_{\text{refl}}}(P) = \frac{1}{N V_P} \sum_{m=0}^{M-1} (C_m - \overline{C})^2.
\]

In the case of rotation, the minimal step between two angles of rotation is \( \Delta \alpha = \frac{2\pi}{N} \). Let \( \alpha_i = i \times \Delta \alpha \). The possible angles of rotation are: \( \alpha_0, \alpha_1, \ldots, \alpha_{N-1} \). The region radial saliency when \( T \) is a rotation is

\[
S_{r_{\text{rot}}}(P) = \frac{1}{N} \sum_{i=0}^{N-1} ncc(P, T_{\alpha_i})
\]

\[
= \frac{1}{N V_P} \left( \sum_{i=0}^{N-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(r_m, \phi_n + \alpha_i)I(r_m, \phi_n) \right) - \frac{1}{M} \left( \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I(r_m, \phi_n) \right)^2
\]

\[
= \frac{1}{N V_P} \left( \sum_{m=0}^{M-1} \left( \sum_{n=0}^{N-1} I(r_m, \phi_n) \right)^2 \right) - \frac{1}{M} \left( \sum_{m=0}^{M-1} \left( \sum_{n=0}^{N-1} I(r_m, \phi_n) \right)^2 \right).
\]
which is the same equation as (5). The radial saliency computed when $T$ is a rotation is equal to the radial saliency when $T$ is a rotation. From now on, we will use only one term, i.e., radial saliency $S_r$ representing both values.

The triple sum of (6) can also be seen as a sum of autocorrelations of discrete signals on circumferences. Similarly, the triple sum of (4) represents a sum of circular convolutions computed for the same discrete signals.

Let us now consider the circumferences as sets of points with intensity values. It is known from statistics that the total sum of squares, in our case computed on the intensity values of $P$, and denoted as $V_P$, can be decomposed into between-sets sum of squares and within-sets sum of squares; in our case into between-circumferences sum of squares and within-circumferences sum of squares. The value of $S_r$ is largest, equal to 1, when $\frac{1}{N} \sum_{n=0}^{N-1} (C_m - C)^2$ or between-circumferences sum of squares is equal to $V_P$; hence, when all variations in $P$ appear between circumferences. Local regions with high values of the between-circumferences sum of squares are blob-like features (see Fig. 12(a)).

Among regions with high within-circumferences sum of squares, we can identify two special types of regions, which are explained in the continuation.

4.1.2 Tangential Saliency

The role of circumferences and radii can be interchanged. By analogy, the tangential saliency $S_t$ is defined and is equal to the normalized between-radii sum of squares:

$$S_t(P) = \frac{1}{N} \frac{1}{V_P} \sum_{n=0}^{N-1} (R_n - R)^2,$$

where $R_n = \sum_{m=0}^{M-1} I(r_m, \phi_n)$ denotes the sum of values on the $n$th radius and $R = \frac{1}{N} \sum_{n=0}^{N-1} R_n$. Then, $S_t$ has the largest value when all changes in a local region appear between radii (see Fig. 12(b)).

4.1.3 Residual Saliency

The third saliency measure is obtained by the help of the dependent effects model, which we have borrowed from statistics, where it is used for analysis of variance [37]. Table 1 shows a decomposition of $P$ by the dependent effects model (see also Fig. 6). In this model, there are three kinds of effects: row effects $\alpha_m$, column effects $\beta_n$, and interaction effects $\gamma_{mn}$. Each element $I_{mn} = I(r_m, \phi_n)$ of $P$ can be represented as

$$I_{mn} = \bar{I} + \alpha_m + \beta_n + \gamma_{mn}.$$

Here, $\bar{I}$ is the mean computed on the intensity values of $P$, and $\alpha_m$ and $\beta_n$ are deviations of the row and column means from $\bar{I}$:

$$\alpha_m = \frac{1}{N} C_m - \bar{I} \quad \text{or} \quad \alpha_m = \frac{1}{N} (C_m - \bar{C}),$$

$$\beta_n = \frac{1}{M} R_n - \bar{I} \quad \text{or} \quad \beta_n = \frac{1}{M} (R_n - \bar{R}).$$

The interaction effect is then:

$$\gamma_{mn} = I_{mn} - (\bar{I} + \alpha_m + \beta_n).$$

The total sum of squares $V_P$ can be expressed as:

$$V_P = V_r + V_t + V_{res} \quad (7)$$

where

$$V_r = N \sum_{m=0}^{M-1} \alpha_m^2, \quad V_t = M \sum_{n=0}^{N-1} \beta_n^2, \quad V_{res} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \gamma_{mn}^2.$$

Dividing (7) by $V_P$, we obtain

$$1 = \frac{V_r}{V_P} + \frac{V_t}{V_P} + \frac{V_{res}}{V_P} = S_r + S_t + S_{res}.$$

The third saliency measure, namely, residual saliency $S_{res}$, can be computed by subtracting $S_r$ and $S_t$ from 1:

$$S_{res} = 1 - S_r - S_t.$$

Fig. 7 summarizes the computation of saliency measures.

4.2 Interest Points

Saliency maps for $S_r$ and $S_t$ are computed over a range of different scales, that is for region with different radii $R$ (see Fig. 8). In a local region, there can be a mixture of symmetries and antisymmetries. For example, points that exhibit local scale-space maxima of $S_r$ represent regions with a locally maximum “amount of symmetry”, while points at local scale-space minima represent regions with a local peak of antisymmetry. Let $\max(lss)$ and $\min(lss)$ denote local scale space maximum and minimum, respectively. Interest points are:

$$\max(lss)(S_r), \quad \max(lss)(S_t),$$

$$\max(lss)(S_{res}) = \max(lss)(1 - S_r - S_t),$$

$$\min(lss)(S_r) = \max(lss)(S_t + S_{res}) = \max(lss)(1 - S_r),$$
TABLE 1

P as Dependent Effects Model

<table>
<thead>
<tr>
<th>$I_0$</th>
<th>$I_{01}$</th>
<th>$I_{0(N-1)}$</th>
<th>$I_{C0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_0 = 7 + \alpha_0 + \beta_0 + \gamma_0$</td>
<td>$I_{01} = 7 + \alpha_0 + \beta_1 + \gamma_0$</td>
<td>$I_{0(N-1)} = 7 + \alpha_0 + \beta_{N-1} + \gamma_0(N-1)$</td>
<td>$I_{C0} = 7 + \alpha_0$</td>
</tr>
<tr>
<td>$I_{10} = 7 + \alpha_1 + \beta_0 + \gamma_1$</td>
<td>$I_{11} = 7 + \alpha_1 + \beta_1 + \gamma_1$</td>
<td>$I_{1(N-1)} = 7 + \alpha_1 + \beta_{N-1} + \gamma_1(N-1)$</td>
<td>$I_{C1} = 7 + \alpha_1$</td>
</tr>
<tr>
<td>$I_{20} = 7 + \alpha_2 + \beta_0 + \gamma_2$</td>
<td>$I_{21} = 7 + \alpha_2 + \beta_1 + \gamma_2$</td>
<td>$I_{2(N-1)} = 7 + \alpha_2 + \beta_{N-1} + \gamma_2(N-1)$</td>
<td>$I_{C2} = 7 + \alpha_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$I_{M-1,0} = 7 + \alpha_{M-1} + \beta_0 + \gamma_{M-1}$</td>
<td>$I_{M-1,1} = 7 + \alpha_{M-1} + \beta_1 + \gamma_{M-1}$</td>
<td>$I_{M-1(N-1)} = 7 + \alpha_{M-1} + \beta_{N-1} + \gamma_{M-1}(N-1)$</td>
<td>$I_{CM-1} = 7 + \alpha_{M-1}$</td>
</tr>
</tbody>
</table>

Fig. 7. Computation of saliency measures.

\[ \min_{lss}(S_t) = \max_{lss}(S_r + S_{res}) = \max_{lss}(1 - S_t), \]
\[ \min_{lss}(S_{res}) = \max_{lss}(S_t + S_r). \]

There are also other measures we can take advantage of, namely, that are $S_r - S_t$, $S_{res} - S_t$, and $S_{res} - S_r$. Additional sets of interest points are indicated by

\[ \max_{lss}(S_r - S_t), \]
\[ \max_{lss}(S_t - S_r), \]
\[ \max_{lss}(S_{res} - S_t) = \max_{lss}(1 - S_r - 2S_t), \]
\[ \max_{lss}(S_t - S_{res}) = \max_{lss}(S_r + 2S_t - 1), \]
\[ \max_{lss}(S_{res} - S_r)) = \max_{lss}(1 - 2S_r - S_t), \]
\[ \max_{lss}(S_r - S_{res}) = \max_{lss}(2S_r + S_t - 1). \]

The interest points represent a rich set of local regions that can be used for image matching and object representation. Figs. 9 and 10 depict a few results.

4.3 Peculiarity of Location

The three saliency measures could also be used for the description of a local region. They represent a point on a unit sphere:

\[ \Upsilon(x, y, R) = [\sqrt{S_r}, \sqrt{S_t}, \sqrt{S_{res}}]. \]

Fig. 11 shows $\Upsilon$ as an RGB colour image. $\Upsilon$ can identify features with different image properties corresponding
Fig. 8. Saliency maps: (from left to right) original image, $S_r$, $S_t$, and $S_{res}$, for the scale parameter (from top to bottom) $R = 7$, $R = 14$, and $R = 28$. The image resolution is for the face image ($228 \times 150$) and for the motorbike image ($262 \times 201$) pixels.

to red, green, and blue color in Fig. 11. Examples of three extreme situations are depicted in Fig. 12. Fig. 12(a) represents examples where $\Upsilon = [1, 0, 0]$ or $S_r = 1$, $S_t = 0$, and $S_{res} = 0$. Here, all changes occur between circumferences. For example, extrema of convex and concave surfaces or blob-like features have large $S_r$ and small $S_t$ and $S_{res}$. Fig. 12(b) represents examples where $\Upsilon = [0, 1, 0]$ or $S_r = 0$, $S_t = 1$, and $S_{res} = 0$. In these examples all changes occur between radii. Points on borders between regions of different intensity values and corner-like features are examples of this kind of points. Fig. 12(c) represents examples where $\Upsilon = [0, 0, 1]$ or $S_r = 0$, $S_t = 0$, and $S_{res} = 1$. In these examples, the sums of values on radii are all equal and the same is true for the sums of values on circumferences. Textured regions have high $S_{res}$ components. Differences between regions as presented in Fig. 12(a), Fig. 12(b), and Fig. 12(c) do not affect the three saliency measures. Hence, we can expect that $\Upsilon$ is robust to within-category variations. $\Upsilon$ is invariant to the group of similarity transformations and to photometric shift. The first property follows from the fact that geometric transformations in the similarity group keep the sets of corresponding points on which normalized correlation coefficients are computed unchanged. The second property, invariance to photometric shift, can be easily seen from equations of the dependent effects model (see Table 1). Effects $\alpha_m$, $\beta_n$, and $\gamma_{mn}$ are independent of the region average value. $\Upsilon$ itself is not powerful enough to be a region descriptor. But, it can easily be noticed that $C_m$ and $R_n$ are nothing other than the zero-th Fourier coefficients of the $m$-th row and $n$-th column of $P$. Higher Fourier coefficients used in a proper way could enrich $\Upsilon$ to become a descriptor.

The proposed approach could also be used for other purposes, for example, as a compression technique. We can do quite a good image reconstruction by using only one type of effects at the interest points. Fig. 13 shows an example of a reconstructed image. By using Fourier coefficients computed on different effects of the dependent effects model, high compression of images could be achieved.

A region of constant intensity value represents a
Fig. 9. A set of local regions corresponding to local scale-space maxima of $S_r$ obtained on 100 images from the Caltech Motorbikes image set with a clear background. The maxima were searched only among the 60 highest local maxima on each image. The maxima were obtained for approximate image resolution $(130 \times 80)$ pixels on a scale from $R = 5$ to $R = 29$.

Fig. 10. Regions belonging to the highest 30 local scale-space maxima of $S_r$ obtained for scales between $R = 5$ and $R = 40$. Image resolution is approximately $140 \times 150$ pixels. Image shows all women from the Caltech Human-Faces image set.

Fig. 11. $\Upsilon$ as an RGB colour image. Red color component corresponds to $S_r$, green to $S_t$, and blue to $S_{\text{res}}$. (From left to right) Original image ($436 \times 228$ pixels); $\Upsilon$ for the scale parameter $R = 7$, $R = 15$, $R = 30$, and $R = 48$. Instead, a look up table of pixels locations is prepared for a local region mask in polar coordinates. The saliency measures are computed by using the following equations for sums of squares:

\[
\sum_{m=0}^{M-1} (C_m - \overline{C})^2 = \sum_{m=0}^{M-1} C_m^2 - \left(\sum_{m=0}^{M-1} C_m / M\right)^2,
\]
\[
\sum_{n=0}^{N-1} (R_n - \overline{R})^2 = \sum_{n=0}^{N-1} R_n^2 - \left(\sum_{n=0}^{N-1} R_n / N\right)^2,
\]
\[
\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{mn} - \overline{I})^2 = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{mn}^2 - \left(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{mn} / MN\right)^2.
\]

5 Implementation Details

The algorithm for interest points detection consists of two main parts:

1) the saliency maps $S_r$ and $S_t$ are first computed for the complete scale-space;
2) scale-space local extrema are determined for each of the saliency measures.

5.1 Computation of $S_r$ and $S_t$

Despite the fact that the proposed approach is developed in polar coordinates, we do not need to transform local regions to polar coordinates to compute $S_r$ and $S_t$. Degenerate case. We have assumed that $\Upsilon = [0,0,0]$ corresponds to a region of constant intensity value.
Fig. 13. Image reconstruction by using the patch mean and only one type of effects of the dependent effects model at the interest points in an image. Different effects are here represented on discs as shown in Fig. 6. The reconstructions are obtained by drawing smaller discs over the larger discs: (from left to right) original image $186 \times 219$ pixels; maxima of $S_t - S_i$ (259 locations); reconstructed image obtained by drawing only row effects at the maxima of $S_t - S_i$; maxima of $S_t$ (228 locations); reconstructed image obtained by drawing only column effects at the maxima of $S_t$.

Here, $I_{mn} = I(r_m, \phi_n)$. Notice that

$$\left( \sum_{m=0}^{M-1} C_m \right)^2 = \left( \sum_{n=0}^{N-1} R_n \right)^2 = \left( \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{mn} \right)^2.$$ 

For a given maximal scale $M$, the algorithm produces $M - 1$ saliency maps for $S_r$ and $S_t$. For each pixel location, the saliency measures are computed through scale in an iterative way where saliency measures at scale $i$ are computed from saliency measures (or their parts) at scale $i - 1$ by adding points of the $i$th circumference. Hence, at each scale $i$, the sums $\sum_{m=1}^{M-1} C_m$, $\sum_{n=1}^{N-1} R_n$, and $\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{mn}$ need to be updated. The most time consuming is the computation of $\sum_{n=0}^{N-1} R_n^2$. It requires all radii at each scale to be updated. For the $n$th radius at the $i$th scale we have:

$$R_n(i) = R_n(i - 1) + I_{in};$$

$$R_n^2(i) =$$

Hence, it takes $MN$ multiplications to compute $\sum_{n=0}^{N-1} R_n^2$ on the complete scale space. Here, $M$ represents the number of circumferences and $N$ the number of radii. By adapting the value $N$ to $M$ ($N = 2\pi M$), time complexity $O(kM^2)$ is achieved with $k$ representing the number of pixels in an image. Compared to Reifeld’s generalized symmetric transform [31], our algorithm computes two complete scale spaces ($2M$ saliency maps) for the same time complexity. It is always possible to speed up the computation by not using all information. One possibility is to use log-polar representation of a local region. The other is to do computations at lower image resolution. In this work, we experiment only with a lower image resolution. We have to think about a parallel approach and special hardware for image processing, as well. This would make it possible to compute the proposed saliency measures in only a few steps.

To use all information available in an image, saliency measures should be computed at subpixel accuracy, e.g., at the pixel center, pixel corners, and mid-points of all four pixel edges. Because this would require a lot of computation, the algorithm computes saliency measures only at the pixel central locations starting with $C_0 = N \times I_{central\_pixel}$ and $R_n(0) = I_{central\_pixel}$, $n = 0, \ldots, N - 1$. A slight smoothing of saliency maps is then applied by replacing the saliency value at a pixel location with the average saliency value computed on the nine neighbouring pixels. This kind of averaging reduces the effects of noise due to discretisation error.

5.2 Computation of Local Extrema

Local maxima and minima for each of the saliency measure are obtained by comparing each pixel value to its eight neighbours in the current saliency map and nine neighbours in the scale level above and below. As proposed in [22], a point is selected only if it has an extreme value compared to all of these neighbours.

6 Experimental Results

In experiments reported here, we have followed two protocols for evaluation of interest points detectors. Twelve self-similarity detectors were tested, namely, local maxima of $S_r$, $S_t$, and $S_{res}$, local minima of $S_r$, $S_t$, and $S_{res}$, and local maxima of $S_r - S_t$, $S_r - S_{res}$, $S_t - S_{res}$, $S_{res} - S_r$, and $S - S_{res}$. Names of detectors are shortened when they are computed on a difference of saliency maps. For example, a difference $S_r - S_t$ is denoted as $S_{r-t}$. In Table 2, local minima detectors are labelled with the letter $a$ (antisymmetry) in front of their names. HS-A, HR-A, ScalSal, MSHar, and DoG stand for the Hessian-Affine detector ([27]), the Harris-Affine detector ([38], [27]), the similarity Saliency detector ([14], the multi-scale Harris detector ([27] without the affine adaptation), and the Difference-Of-Gaussian blob detector ([22], respectively. Due to readability of the graphs, only the results for four self-similarity detectors are shown in graphs (see Figs. 14, 19-24, 26, and 27).

6.1 Intraclass Variations and Image Perturbations

The aim here is to measure the performance of self-similarity detectors under intra-class variations and image perturbations in the same way as proposed in [15]. The detectors were tested on 450 images from the Caltech Human-Faces set, 200 images from
Fig. 14. (a) Repeatability results for the Caltech Human-Faces images. (b) Repeatability results for the Caltech Cars images. (c) Repeatability results for the Caltech Motorbikes images with a clear background. (d) Repeatability results for the Caltech Motorbikes images with background clutter.

Fig. 15. The 60 best local maxima of $S_r$, obtained for the same person on two different backgrounds. Due to the regular structure in the background area in the image on the right, many of the best local maxima belong to the background area, and hence, fewer matches are found on the face.

the Caltech Motorbikes set¹ (84 images have clear background and 116 images have clutter in the background), and all 126 Caltech Cars (Rear 2) images. Affine transformations between objects of interest in images were estimated by using the ground-truth locations given at http://www.robots.ox.ac.uk/~timork/Saliency/AffineInvariantSaliency.html. Tests were performed on images with lower image resolution². Scores were computed on a scale from $R = 2$ to $R = 24$ of downsampled images. We consider that a region matches if it fulfills the three requirements recommended in [15] with the following conditions: its position matches within 10 pixels of the original image size; its scale is within 20% and the normalized mutual information between the appearances $(MI(A, B) = 2(H(A) + H(B) - H(A, B))/(H(A) + H(B)))$ is greater than 0.2. The average correspondence score $S$ proposed in [15] is measured as follows. The $N$ best regions are detected in each of the $M$ images in the dataset. For a particular reference image $i$, the correspondence score $S_i$ is given by the proportion of correspondences with detected regions for all other images in the dataset, i.e.,

$$S_i = \frac{\text{Total number of matches}}{\text{Total number of detected regions}} = \frac{N_i}{M}.$$

$S_i$ is computed for $\frac{N_i}{M}$ different selections and averaged. Fig. 14 shows the results obtained.

On the Human Faces images, the best results (Fig. 14(a)) are obtained for the local scale-space maxima of $S_r$ and $S_r - S_t$. The local scale-space maxima of $S_t$ and $S_{res}$ give low scores. The local scale-space maxima of $S_t$ represent features at the borders between regions of different intensity and corner-like features. The local scale-space maxima can trace along straight borders. When this happens on the background, they do not have corresponding matches in other images. Corner-like features are points with high curvature on region and object borders, so we can expect larger variations in their position and appearance due to background, compared to region centers. The best local scale-space maxima of $S_{res}$ are for the Faces images often features of the background area. The low average correspondence scores obtained do not mean that there are no local scale-space maxima in the area of the faces or that they are not repeatable there. They are not only among the highest 60 local scale-space maxima. The same can be true for other features too. The results obtained depend on the background area. Fig. 15 shows an example of this.

¹ This image set is a part of the Caltech-101 object categories database.

² Images were down sampled by replacing pixel value at location $(i, j)$ with $\frac{1}{4}(I(i, j) + \frac{1}{2}(I(i-1, j) + I(i-1, j+1) + I(i+1, j) + I(i+1, j+1)))$ and taking only pixels with even $i$ and $j$. Images of the Faces and Cars sets were down sampled two times. Images of the Motorbikes set were down sampled only once.
notice that the highest local scale-space maxima of $S_r$ can also be found in the background when there is a regular structure. In such cases, fewer matches are found on the area of face.

Fig. 14(b)) shows the results obtained on the Caltech-Cars image set. The similarity Saliency region detector obtained for this image set the best scores. The lower average correspondence scores obtained for the Caltech-Cars image set are due to larger variability of car parts. Locations of parts like number plates, exhausts, front mirrors, back bumpers, and car lights vary and cannot be transformed from one image to another by an affine transformation (see Fig. 16).

The results obtained on the Motorbikes image set (Figs. 14(c) and 14(d)) are better than the results obtained on the Cars and Faces image sets. This is partly due to lower image resolution of the Motorbike images, meaning that position matches within 10 pixels can more easily be accomplished. The performance of the local scale-space maxima of $S_r$ and $S_{res}$ is here different than on the Faces and Cars image sets. To understand the obtained results better, we show in Fig. 17 two examples of images with marked local regions obtained for the local scale-space maxima of $S_r$, $S_t$, and $S_{res}$. One image is from the set of images with a clear background. The other is from the set of images with clutter in the background. The local scale-space maxima of $S_{res}$ represent large regions. In the case of the clear background, they appear mostly on the background and represent white regions with only a few dark points on the border. JPEG compression moves circles away from the motorbike border. These regions are similar from image to image and are highly repeatable. In the case of background clutter, the highest local scale-space maxima of $S_{res}$ can be found in the area of the motorbikes. They are repeatable and give high average correspondence scores. The border between the motorbike and the clear background represents locations with high $S_t$. Most of the local scale-space maxima of $S_t$ are obtained on that border and fewer inside the area of the motorbike, which makes features highly repeatable. On images with clutter, $S_t$ for locations on the border is reduced due to clutter (see also Fig 8). The local scale-space maxima of $S_t$ can now be higher on the background, for example, on the border of the shadow on the floor. These locations are not repeatable; therefore, the local scale-space maxima of $S_t$ obtained on images with clutter give lower average correspondence scores. There are also other reasons that make the results on images with clutter worse: different lighting conditions as pictures were taken outside, occlusions by other objects, different viewing directions, and different orientations of some parts of the motorbike, such as the front wheel. We also feel that the images with clutter exhibit greater variability among the different motorbike images. On the image set with clear background, the local maxima of $S_{res}$ obtained the best scores. The scores obtained by
the local scale-space maxima of $S_r - S_t$ are also high. On the image set with background clutter, the similarity Saliency region detector obtained the best results.

### 6.2 Varying Blur, JPEG Artifacts, Lighting Change, In-Plane Orientation, Scale, and Viewpoint Change

The aim of the experiments reported here is to test the self-similarity detectors in accordance with the protocol suggested in [26]. Detectors are compared on the basis of the number of corresponding regions detected in images under different geometric and photometric transformations. Fig. 18 shows images from the test sets. The reference image is the left-most image of each set. First, accuracy and repeatability of the detection are tested. The ground truth in all cases is provided by mapping the local regions detected on the test image to a reference image using homographies. The basic measure of accuracy is the relative amount of overlap between corresponding regions. Results reported in Figs. 19-27 are obtained for 40% overlap error, while results are given in Table 2 for 20% overlap error. Two measures are reported: the repeatability score and the absolute number of corresponding regions found in the test image. The repeatability score for a given pair of images is computed as the ratio between the number of region-to-region correspondences and the smaller of the number of regions obtained in the pair of images. Second, the distinctiveness of the detected region is tested. Detected regions are represented by the SIFT descriptor [22]. Two measures are reported: the matching score and the number of correct matches. The matching score is the ratio between the number of correct matches and the smaller number of detected regions in the pair of images. A match is the nearest neighbour in the descriptor space (euclidian distance). A match is deemed correct if the overlap error of corresponding regions is less than 40%.

In all experiments reported here, interest points were computed on images with lower image resolution. Nine neighbouring pixels of the original image were averaged and represented as one pixel. All local extrema for scales from $R = 4$ to $R = 24$ (from $R = 12$ to $R = 69$ for the original image resolution) were taken into account and no threshold was used. The obtained coordinates and radii of the detected regions were then transformed to the original image sizes, and all further computations were done by using the original homographies and software from http://www.robots.ox.ac.uk/~vgg/research/affine.

#### 6.2.1 Blur

Figs. 19 and 20 show the results obtained for the structured and textured scene, both undergoing increasing amounts of image blur (see Figs. 18(a) and 18(b)). The best results are obtained for the local scale-space maxima of $S_r - S_t$ for both scene types. We see two reasons for this. The first reason can be clarified by Fig. 12(a). The topmost example in Fig. 12(a) is represented with the same values of $S_r$, $S_t$, and $S_{res}$ as a black disc with a white ring around it. Hence, we can expect that the saliency measures of blob-like features are not much affected by blur. The second reason is more general and can also be valid for other types of image transformations. Let transformation $T$, applied on an image, change the three saliency measures $S_r$, $S_t$, and $S_{res}$ at particular location for $\Delta_r$, $\Delta_t$, and $\Delta_{res}$, respectively. The value of $S_r - S_t$ changes for $\Delta_r - \Delta_t$. When the changes $\Delta_r$ and $\Delta_t$ are similar, the difference $\Delta_r - \Delta_t$ is small, and hence, the value of $S_r - S_t$ is less affected by $T$ than the values of $S_r$, $S_t$, and $S_{res}$. This, of course, is not true for all points in an image and the difference $\Delta_r - \Delta_t$ can also be larger than $\Delta_r$, $\Delta_t$ and $\Delta_{res}$. But high scores obtained for the maxima of $S_r - S_t$ suggest that, at the local scale-space maxima of $S_r - S_t$, this might be the case.
The numbers of corresponding regions detected on the structured scene are lower than they are on the textured scene for most detectors. Exceptions are the local scale-space maxima of $S_t$. The matching scores obtained by self-similarity detectors are high compared to the matching scores obtained by the Hessian-Affine, Harris-Affine, and MSER detector. For 20% overlap error (Table 2) the local scale-space maxima of $S_{res} - S_t$ obtained the best average repeatability score on the structured scene.

6.2.2 JPEG Artifacts

Fig. 21 shows the repeatability scores for the JPEG compression sequence from Fig. 18(c). The Hessian-Affine and Harris-Affine detector show here the best performance. The local scale-space maxima of $S_r - S_t$ gave the highest repeatability scores among self-similarity detectors.

6.2.3 Lighting Changes (Fig. 18(d))

Results for lighting changes are shown in Fig. 22. The best repeatability scores are obtained by the MSER detector. The local scale-space maxima of $S_r - S_t$ and $S_r$ give slightly lower repeatability scores. But the number of correspondences obtained by the local scale-space maxima of $S_r - S_t$ are approximately three times larger than the number of correspondences obtained by the MSER detector. The matching scores of self-similarity detectors and the MSER detector are high compared to the matching scores obtained by the Hessian-Affine and Harris-Affine detector.
6.2.4 Scale Change and In-Plane Rotation

Results for scale changes and in-plane rotations are shown in Figs. 23 and 24. Features in all test images are detected for fixed scale range. Because features get smaller from image to image, the relative scale range is different for each test image. This makes results difficult to interpret. On the structured scene, the local scale-space maxima of $S_{r,s}$ give the best repeatability scores (Fig. 23). The local scale-space maxima of $S_{r,s}$ have large percentage of regions with radii of middle size (see Fig. 28), and hence, fewer regions are lost due to the fixed scale range. On the textured scene, the best repeatability scores are obtained by the Hessian-Affine detector. Notice that the number of correspondences is small. For 20% overlap error (Table 2) the best average repeatability score is obtained by the local scale-space maxima of $S_r$.

6.2.5 Viewpoint Change

Local regions obtained by our detectors are discs (Fig. 25), hence, they are not adapted to change in viewing angle. Our intention here has been to see how far we can go with the proposed approach. Results for the structured scene (Fig 18(g)) are shown on Fig. 26. The MSER detector gives here the best results. Still, for 20° viewing angle change, the local scale-space maxima of $S_r - S_t$ obtained slightly higher repeatability score than the MSER detector. For a 30° viewpoint angle change, the repeatability scores of self-similarity detectors are higher than the repeatability score obtained by the Harris-Affine detector. For a 50° viewpoint angle change, no correspondence is detected that fulfils the default setting,
Fig. 25. The 120 highest maxima of $S_r - S_t$ obtained on the first and the third image of the Graffiti sequence (Fig. 18(g)).

Fig. 26. Viewpoint change for the structured scene (Graffiti sequence Fig. 18(g)). (a) Repeatability score for viewpoint change. (b) Number of corresponding regions. (c) Matching score. (d) Number of correct nearest matches.

Fig. 27. Viewpoint change for the textured scene (Wall sequence Fig. 18(h)). (a) Repeatability score for viewpoint change. (b) Number of corresponding regions. (c) Matching score. (d) Number of correct nearest matches.

i.e., 40% overlap error. On the textured scene (Fig. 18(h)), self-similarity detectors have obtained surprisingly good results (see Fig. 27). The local scale-space maxima of $S_r - S_t$, give the best repeatability scores up to $50^\circ$.

6.2.6 Region Size

Figure 28 shows histograms of region size for different self-similarity detectors. The regions belonging to the local scale-space maxima of $S_{res}$ tend to have larger radii than the local scale-space maxima of $S_r - S_t, S_r, and S_t$. We can also notice that the local scale-space maxima of $S_r - S_t$ have larger percentage of regions with radii of middle size than the local scale-space maxima of $S_r$ and $S_t$.

7 CONCLUSION

In this paper we have presented a new approach to detecting interest points in images. The total sum of squares computed on intensity values of a local region is divided into three components: between circumferences sum of squares, between radii sum of squares, and the remainder. The three components normalized by the total sum of squares determine three new region saliency measures and are computed at different scales. Here scale corresponds to the radius of a local region. Local extrema in scale-space are located for each of the saliency measures. The extrema are features with complementary image properties.

The performance of the new approach was demon-
stratified on a wide variety of image sets. The obtained results compare favourably with the results obtained by the leading interest point detectors from the literature. The proposed approach gives a rich set of highly distinctive local regions that can be used for object recognition and image matching.

Our future research goal is to extend the proposed approach to a local region descriptor by using higher Fourier coefficients computed on different effects of the dependent effects model. Next, we would like to find criteria for joining local regions into larger structures and then simplify these structures into primitives that can be used for object representation and recognition.

A deficiency of the proposed method is that it does not include an affine adaptation of local regions. Hence, the proposed method is not appropriate for tasks like wide based stereo. However, this is not a fundamental limitation of the method. Circumferences can be replaced with ellipses and this is one of the possible directions for the future work.

**APPENDIX**

Rotation $T$ with $\alpha = \frac{2\pi}{n}$ maps locations $(r, \alpha + \phi)$ to $(r, (i+1)\alpha + \phi)$; $i = 0, 1, \ldots, n - 1$. Because $I(r, (i+1)\alpha + \phi) = a + bI(r, \alpha + \phi)$ and $(r, na + \phi) = (r, \phi)$ we can express the relation between $I(r, \alpha + \phi)$ and $I(r, \phi)$ in two different ways:

$$I(r, \alpha + \phi) = a + bI(r, \phi)$$

$$I(r, \phi) = \sum_{i=0}^{n-1} ab^i + b^{n-1}I(r, \alpha + \phi). \quad (8)$$

By replacing variable $I(r, \phi)$ in the first line of (8) with $\sum_{i=0}^{n-1} ab^i + b^{n-1}I(r, \alpha + \phi)$ and in the second line $I(r, \alpha + \phi)$ with $a + bI(r, \phi)$ we obtain:

$$I(r, \alpha + \phi) = \sum_{i=0}^{n-1} ab^i + b^ni(r, \alpha + \phi)$$

$$I(r, \phi) = \sum_{i=0}^{n-1} ab^i + b^nI(r, \phi). \quad (9)$$

Let us now subtract the second equation of (9) from the first:

$$I(r, \alpha + \phi) - I(r, \phi) = b^nI(r, \alpha + \phi) - I(r, \phi)). \quad (10)$$

From here, $b^n = 1$ and $\sum_{i=0}^{n-1} ab^i = 0$. For even $n$, two solutions for $b$ are $b = 1$ and $b = -1$. Solution $b = 1$ gives $\sum_{i=0}^{n-1} ab^i = na$ and from here $a = 0$ while solution $b = -1$ gives $\sum_{i=0}^{n-1} ab^i = 0$. From the first line of (8) it follows that $a = I(r, \phi) + I(r, \alpha + \phi)$. For odd $n$, parameter $b$ can only have one value, that is $b = 1$. In this case $\sum_{i=0}^{n-1} ab^i = na$ and hence $a = 0$.

**ACKNOWLEDGMENT**

The author would like to thank Richard I. Hartley, Jiří Matas, and Jan Šochman for reading the paper draft and giving valuable comments, and Michal Perd'och, Dušan Řomerčević, and Luka Čehovin for help with software and experimental work. This research work has been supported by the Research program Computer Vision P2-0214 (RS).

**REFERENCES**


